

Equivalence of two orthogonalities between probability measures

Asuka Takatsu*

Abstract

Given any two probability measures on a Euclidean space with mean 0 and finite variance, we demonstrate that the two probability measures are orthogonal in the sense of Wasserstein geometry if and only if the two spaces by spanned by the supports of each probability measure are orthogonal.

1 Introduction and Main theorem

This paper is concerned with the two orthogonalities between a pair of probability measures on \mathbb{R}^d : one is measured by the Wasserstein metric and the other is given in terms of the orthogonality between spaces spanned by the supports of probability measures.

Let \mathcal{P}_2 be the set of Borel probability measures μ on \mathbb{R}^d with finite variance, namely

$$\int_{\mathbb{R}^d} |x|^2 d\mu(x) < \infty.$$

Given any $\mu, \nu \in \mathcal{P}_2$, we define their (L^2 -)Wasserstein distance by

$$W_2(\mu, \nu) := \sqrt{\inf_{\sigma \in \Pi(\mu, \nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 d\sigma(x, y)}, \quad (1.1)$$

where $\Pi(\mu, \nu)$ is the set of probability measures σ on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals μ and ν , that is, $\sigma[B \times \mathbb{R}^d] = \mu[B]$ and $\sigma[\mathbb{R}^d \times B] = \nu[B]$ hold for all Borel sets $B \subset \mathbb{R}^d$. The pair (\mathcal{P}_2, W_2) is a metric space and inherits several properties of \mathbb{R}^d (for instance see [Vi, Section 6]). For example, (\mathcal{P}_2, W_2) has a cone structure as well as \mathbb{R}^d . We say that a metric space (X, d_X) has a cone structure if there exists a metric space (Σ, \angle) such that (X, d_X) is isometric to the quotient space $(\Sigma \times [0, \infty)/ \sim, d_C)$, where the equivalence relation \sim is defined by $(\xi, s) \sim (\eta, t)$ if we have $(\xi, s) = (\eta, t)$ or $s = t = 0$, and the distance function d_C is given by

$$d_C((\xi, s), (\eta, t)) := \sqrt{s^2 + t^2 - 2st \cos(\min\{\angle(\xi, \eta), \pi\})}.$$

*Graduate School of Mathematics, Nagoya University, Nagoya 464-8602, Japan (takatsu@math.nagoya-u.ac.jp) & Max-Planck-Institut für Mathematik, Vivatsgasse 7, 53111 Bonn, Germany.

Mathematics Subject Classification (2010): 60D05; 51F20.

keywords: Wasserstein geometry; orthogonality.

Of course, \mathbb{R}^d is isometric to the quotient space of $(d-1)$ -sphere with its standard metric and it was proved in [TY] that (\mathcal{P}_2, W_2) is isometric to the quotient space of

$$\mathcal{P}_2 := \{\mu \in \mathcal{P}_2 \mid W_2(\delta_0, \mu) = 1\}$$

with the metric given by

$$\angle(\mu, \nu) := \arccos \left(1 - \frac{1}{2} W_2(\mu, \nu)^2 \right).$$

We remark that $\angle(\mu, \nu)$ is regarded as the angle between the two Wasserstein geodesics from δ_0 to μ and from δ_0 to ν , which in addition coincides with the comparison angle of $\angle_{\mu\delta_0\nu}$ defined by the cosine formula of the form

$$\angle(\mu, \nu) = \angle_{\mu\delta_0\nu} := \arccos \left(\frac{W_2(\delta_0, \mu)^2 + W_2(\delta_0, \nu)^2 - W_2(\mu, \nu)^2}{2W_2(\delta_0, \mu)W_2(\delta_0, \nu)} \right).$$

Note that this formula can be extended to any pair in $\mathcal{P}_2 \setminus \{\delta_0\}$. Moreover, \mathcal{P}_2 is isometric to the direct product of \mathbb{R}^d and the convex subspace given by

$$\mathcal{P}_{2,0} := \left\{ \mu \in \mathcal{P}_2 \mid \int_{\mathbb{R}^d} x d\mu(x) = 0 \right\}$$

which also has a cone structure and contains no line, that is its vertex angle given by

$$\sup \{ \angle(\mu, \nu) \mid \mu, \nu \in \mathcal{P}_{2,0} \setminus \{\delta_0\} \}$$

is less than π . Indeed for any $\mu, \nu \in \mathcal{P}_{2,0} \setminus \{\delta_0\}$, the marginals of the product measure $\mu \times \nu$ are obviously μ and ν , and the fact

$$W_2(\delta_0, \mu)^2 = \int_{\mathbb{R}^d} |x|^2 d\mu(x), \quad W_2(\delta_0, \nu)^2 = \int_{\mathbb{R}^d} |y|^2 d\nu(y)$$

together with the condition that the means of μ and ν are 0 yield that

$$W_2(\mu, \nu)^2 \leq \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 d(\mu \times \nu)(x, y) = W_2(\delta_0, \mu)^2 + W_2(\delta_0, \nu)^2,$$

which shows

$$\angle(\mu, \nu) = \arccos \left(\frac{W_2(\delta_0, \mu)^2 + W_2(\delta_0, \nu)^2 - W_2(\mu, \nu)^2}{2W_2(\delta_0, \mu)W_2(\delta_0, \nu)} \right) \leq \frac{\pi}{2}. \quad (1.2)$$

In the case of $d \geq 2$, the equality in (1.2) holds, for example, by taking $\mu := (\delta_\xi + \delta_{-\xi})/2$ and $\nu := (\delta_\eta + \delta_{-\eta})/2$ with $\xi, \eta \in \mathbb{R}^d \setminus \{0\}$ satisfying that $\langle \xi, \eta \rangle = 0$.

We provide a necessary and sufficient condition for a pair in $\mathcal{P}_{2,0} \setminus \{\delta_0\}$ to attain the equality in (1.2), in other words, for a pair of Wasserstein geodesics in $\mathcal{P}_{2,0}$ starting from δ_0 to be orthogonal. For $A \subset \mathbb{R}^d$, let $\text{Span}\langle A \rangle$ denote the linear span of A .

Theorem *Given any $\mu, \nu \in \mathcal{P}_{2,0} \setminus \{\delta_0\}$, the condition $\angle(\mu, \nu) = \pi/2$ is equivalent to the orthogonality between $\text{Span}\langle \text{supp}(\mu) \rangle$ and $\text{Span}\langle \text{supp}(\nu) \rangle$.*

Acknowledgements. The author would like to express her gratitude to Cédric Villani for his valuable comments and discussions. She is grateful for the hospitality of Institut des Hautes Études Scientifiques, where most of this work was done. She would also like to thank Takumi Yokota for his comments. This work is partially supported by JSPS-IHÉS (EPDI) fellowship.

2 Proof of Theorem

We first prepare two lemmas derived from the feature of mean.

Lemma 2.1 *For any $\mu \in \mathcal{P}_{2,0} \setminus \{\delta_0\}$, there exist $\{\xi_j\}_{j=0}^k \subset \text{supp}(\mu)$ and $\{a_j\}_{j=0}^k \subset \mathbb{R}$ such that $\sum_{j=0}^k a_j \xi_j = 0$ and $\sum_{j=0}^k a_j \neq 0$.*

Proof. Suppose that the dimension of $\text{Span}\langle \text{supp}(\mu) \rangle$ equals k and $\{\xi_i\}_{i=1}^k \subset \text{supp}(\mu)$ is a basis of $\text{Span}\langle \text{supp}(\mu) \rangle$. For any orthonormal basis $\{u_i\}_{i=1}^k$ of $\text{Span}\langle \text{supp}(\mu) \rangle$, there exists $\{a_{ij}\}_{1 \leq i,j \leq k} \subset \mathbb{R}$ such that $u_i = \sum_{j=1}^k a_{ij} \xi_j$. We define the functions p_i and a_j on \mathbb{R}^d by $p_i(x) := \langle x, u_i \rangle$ and $a_j(x) := \sum_{i=1}^k p_i(x) a_{ij}$. If $\xi_0 \in \text{supp}(\mu)$ satisfies $\sum_{j=1}^k a_j(\xi_0) \neq 1$, then $\{\xi_j\}_{j=0}^k$ and $\{a_j(\xi_0)\}_{j=0}^k$ with $a_0(\xi_0) = -1$ are the desired families since we have

$$\xi_0 = \sum_{i=1}^k p_i(\xi_0) u_i = \sum_{i=1}^k p_i(\xi_0) \left(\sum_{j=1}^k a_{ij} \xi_j \right) = \sum_{j=1}^k \left(\sum_{i=1}^k p_i(\xi_0) a_{ij} \right) \xi_j = \sum_{j=1}^k a_j(\xi_0) \xi_j.$$

Such a point always exists, otherwise any $x \in \text{supp}(\mu)$ satisfies $\sum_{j=1}^k a_j(x) = 1$ and we have a contradiction as

$$1 = \int_{\mathbb{R}^d} \sum_{j=1}^k a_j(x) d\mu(x) = \int_{\mathbb{R}^d} \sum_{j=1}^k \sum_{i=1}^k p_i(x) a_{ij} d\mu(x) = \int_{\mathbb{R}^d} \left\langle x, \sum_{j=1}^k \sum_{i=1}^k a_{ij} u_i \right\rangle d\mu(x) = 0,$$

where the last equality follows from the definition of mean. \square

For a point $\xi \in \mathbb{R}^d$ and a family $\{\Xi_\lambda\}_{\lambda \in \Lambda}$ of subsets in \mathbb{R}^d , we set

$$\Xi_\lambda - \xi := \{\xi_\lambda - \xi \mid \xi_\lambda \in \Xi_\lambda\}, \quad \sum_{\lambda \in \Lambda} \Xi_\lambda := \left\{ \sum_{i=1}^n \xi_{\lambda_i} \mid \xi_{\lambda_i} \in \Xi_{\lambda_i}, \lambda_i \in \Lambda, \exists n \in \mathbb{N} \right\}.$$

Lemma 2.2 *Given any $\mu \in \mathcal{P}_{2,0} \setminus \{\delta_0\}$, we have*

$$\sum_{\xi \in \text{supp}(\mu)} \text{Span}\langle \text{supp}(\mu) - \xi \rangle = \text{Span}\langle \text{supp}(\mu) \rangle.$$

Proof. Since the relation $\text{Span}\langle \text{supp}(\mu) - \xi \rangle \subset \text{Span}\langle \text{supp}(\mu) \rangle$ is trivially true for any $\xi \in \text{supp}(\mu)$, the relation $\sum_{\xi \in \text{supp}(\mu)} \text{Span}\langle \text{supp}(\mu) - \xi \rangle \subset \text{Span}\langle \text{supp}(\mu) \rangle$ is also true.

Lemma 2.1 ensures the existences of $\{\xi_j\}_{j=0}^k \subset \text{supp}(\mu)$ and $\{a_j\}_{j=0}^k \subset \mathbb{R}$ such that $\sum_{j=0}^k a_j \xi_j = 0$ and $a := \sum_{j=0}^k a_j \neq 0$. Then for any $x \in \text{Span}\langle \text{supp}(\mu) \rangle$, we find that

$$x = \sum_{j=0}^k \frac{a_j}{a} (x - \xi_j) \in \sum_{\xi \in \text{supp}(\mu)} \text{Span}\langle \text{supp}(\mu) - \xi \rangle.$$

\square

Let us now prove Theorem by using the following known result.

Lemma 2.3 ([Vi, Theorem 5.10(ii)]) *Given any $\mu, \nu \in \mathcal{P}_2$ and any $\sigma \in \Pi(\mu, \nu)$, the following two properties are equivalent to each other.*

(i) σ attains the infimum in (1.1).

(ii) For any $n \in \mathbb{N}$ and $\{(x_i, y_i)\}_{i=1}^n \subset \text{supp}(\sigma)$ with the convention $y_{n+1} = y_1$, we have

$$\sum_{i=1}^n |x_i - y_i|^2 \leq \sum_{i=1}^n |x_i - y_{i+1}|^2.$$

Proof of Theorem. Take any $\mu, \nu \in \mathcal{P}_{2,0} \setminus \{\delta_0\}$ and fix them. We first remark that $\mu \times \nu$ attains the infimum in (1.1) if and only if $W_2(\mu, \nu)^2 = W_2(\delta_0, \mu)^2 + W_2(\delta_0, \nu)^2$ holds.

If $\angle(\mu, \nu) = \pi/2$, then $\mu \times \nu$ attains the infimum in (1.1) due to (1.2). For any $(x, y), (\xi, \eta) \in \text{supp}(\mu \times \nu)$, $(x, \eta), (\xi, y)$ also lie in $\text{supp}(\mu \times \nu)$ and Lemma 2.3 yields that

$$|x - y|^2 + |\xi - \eta|^2 \leq |x - \eta|^2 + |\xi - y|^2 \leq |x - y|^2 + |\xi - \eta|^2$$

meaning $\langle x - \xi, y - \eta \rangle = 0$. Since we take $(x, y) \in \text{supp}(\mu \times \nu)$ arbitrarily, the spaces $\text{Span}\langle \text{supp}(\mu) - \xi \rangle$ and $\text{Span}\langle \text{supp}(\nu) - \eta \rangle$ are orthogonal. Moreover, Lemma 2.2 with arbitrary choice of $(\xi, \eta) \in \text{supp}(\mu \times \nu)$ provides the orthogonality between $\text{Span}\langle \text{supp}(\mu) \rangle$ and $\text{Span}\langle \text{supp}(\nu) \rangle$.

Conversely suppose the orthogonality between $\text{Span}\langle \text{supp}(\mu) \rangle$ and $\text{Span}\langle \text{supp}(\nu) \rangle$. Then $\mu \times \nu \in \Pi(\mu, \nu)$ satisfies condition (ii) in Lemma 2.3 and thus $\mu \times \nu$ attains the infimum in (1.1), which in turn implies

$$\angle(\mu, \nu) = \arccos \left(\frac{W_2(\delta_0, \mu)^2 + W_2(\delta_0, \nu)^2 - W_2(\mu, \nu)^2}{2W_2(\delta_0, \mu)W_2(\delta_0, \nu)} \right) = \frac{\pi}{2}.$$

□

Remark 2.4 (1) The “if” part can be proved in a different way: for any $\mu, \nu \in \mathcal{P}_2 \setminus \{\delta_0\}$, let $\theta(\mu, \nu)$ be the smallest principal angle between $\text{Span}\langle \text{supp}(\mu) \rangle$ and $\text{Span}\langle \text{supp}(\nu) \rangle$, that is,

$$\theta(\mu, \nu) := \min \left\{ \arccos \frac{\langle x, y \rangle}{|x||y|} \mid x \in \text{Span}\langle \text{supp}(\mu) \rangle, y \in \text{Span}\langle \text{supp}(\nu) \rangle, x, y \neq 0 \right\}.$$

Note that $\theta(\mu, \nu) \in [0, \pi/2]$. If the relation $\angle(\mu, \nu) \geq \theta(\mu, \nu)$ holds, then the orthogonality between $\text{Span}\langle \text{supp}(\mu) \rangle$ and $\text{Span}\langle \text{supp}(\nu) \rangle$, that is $\theta(\mu, \nu) = \pi/2$, implies $\angle(\mu, \nu) = \pi/2$.

To prove the relation $\angle(\mu, \nu) \geq \theta(\mu, \nu)$, let $\sigma \in \Pi(\mu, \nu)$ be optimal. Then σ is supported on $\text{Span}\langle \text{supp}(\mu) \rangle \times \text{Span}\langle \text{supp}(\nu) \rangle$ and Hölder’s inequality yields that

$$\begin{aligned} W_2(\mu, \nu)^2 &\geq W_2(\delta_0, \mu)^2 + W_2(\delta_0, \nu)^2 - 2 \cos \theta(\mu, \nu) \int_{\mathbb{R}^d \times \mathbb{R}^d} |x||y| d\pi(x, y) \\ &\geq W_2(\delta_0, \mu)^2 + W_2(\delta_0, \nu)^2 - 2 \cos \theta(\mu, \nu) W_2(\delta_0, \mu) W_2(\delta_0, \nu), \end{aligned} \quad (2.1)$$

which shows $\angle(\mu, \nu) \geq \theta(\mu, \nu)$ as desired. The relation $\angle(\mu, \nu) = \theta(\mu, \nu)$ holds if and only if all the inequalities in (2.1) are equalities, which is equivalent to the existence of $a \in \mathbb{R}$ such that $\langle x, y \rangle = a|x|^2 \cos \theta(\mu, \nu)$ for σ -a.e. $(x, y) \in \mathbb{R}^d \times \mathbb{R}^d$.

(2) In the case of $d = 1$, Theorem yields that the angle between any pair in $\mathcal{P}_{2,0} \setminus \{\delta_0\}$ is strictly less than $\pi/2$, however the vertex angle equals $\pi/2$. Indeed, set

$$\begin{aligned}\mu &:= \frac{1}{2} \{\delta_1 + \delta_{-1}\}, \quad \nu_n := \left(1 - \frac{1}{n^2}\right) \delta_0 + \frac{1}{2n^2} \{\delta_n + \delta_{-n}\}, \\ \sigma_n &:= \frac{1}{2} \left(1 - \frac{1}{n^2}\right) \{\delta_{(1,0)} + \delta_{(-1,0)}\} + \frac{1}{2n^2} \{\delta_{(1,n)} + \delta_{(-1,-n)}\}\end{aligned}$$

for any $n \in \mathbb{N}$. Then we have $\mu, \nu_n \in \mathcal{P}_{2,0}$ and $\sigma_n \in \Pi(\mu, \nu_n)$ attains the infimum in (1.1), which implies

$$\angle(\mu, \nu_n) = \arccos \left(\frac{W_2(\delta_0, \mu)^2 + W_2(\delta_0, \nu)^2 - W_2(\mu, \nu)^2}{2W_2(\delta_0, \mu)W_2(\delta_0, \nu)} \right) = \arccos \frac{1}{n}.$$

References

- [TY] A. Takatsu and T. Yokota, Cone structure of L^2 -Wasserstein spaces, preprint (2010). Available at arXiv:0812.2752
- [Vi] C. Villani, Optimal transport, old and new, Springer-Verlag, Berlin, 2009.